(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
26 October 2006 (26.10.2006)

PCT

(10) International Publication Number
WO 2006/112793 A1

(51) International Patent Classification:
G11C 11/34 (2006.01)

(21) International Application Number:
PCT/SG2005/000129

(22) International Filing Date: 20 April 2005 (20.04.2005)

(25) Filing Language: English

(26) Publication Language: English

(71) Applicant *(for all designated States except US)*: **NATIONAL UNIVERSITY OF SINGAPORE** [SG/SG]; 10 Kent Ridge Crescent, Singapore 119260 (SG).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: **SAMANTA, Santanu, Kumar** [IN/SG]; National University of Singapore, 10 Kent Ridge Crescent, Singapore 119260 (SG). **YOO, Won, Jong** [US/SG]; National University Of Singapore, 10 Kent Ridge Crescent, Singapore 119260 (SG).

(74) Agent: **VIERING, JENTSCHURA & PARTNER**; P.O. Box 1088, Rochor Post Office, Rochor Road, Singapore 911833 (SG).

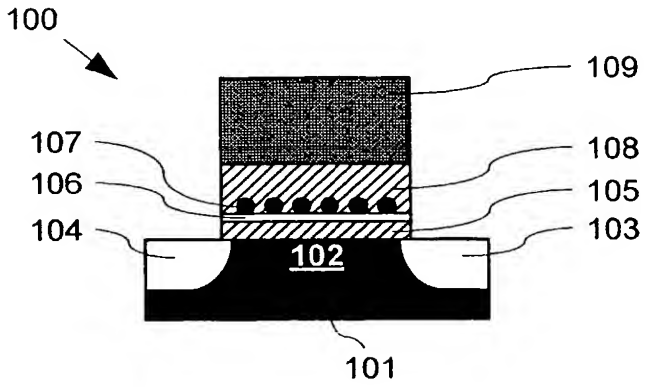
(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SM, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: NONVOLATILE FLASH MEMORY DEVICE AND METHOD FOR PRODUCING THE SAME



(57) Abstract: A method of producing metallic nanocrystals (107) embedded in high-k dielectric material as well as a nonvolatile flash memory device (100) comprising a discrete charge carrier storage layer, the discrete charge carrier storage layer comprising metallic nanocrystals (107) embedded in high-k dielectric material. In the method described in this invention, firstly an ultra-thin metal film is deposited over a first (105) and a second (106) dielectric layer including high-k dielectric material provided on a substrate (101). Then, the ultra-thin metal film is annealed for forming the metallic nanocrystals (107) on the second dielectric layer (106). Finally, the second dielectric layer (106) and the metallic nanocrystals (107) are covered with a third dielectric layer (108) of high-k dielectric material for forming metallic nanocrystals (107) embedded in high-k dielectric material. The first (105), second (106) and third (108) dielectric layers together with the embedded metallic nanocrystals (107) as the discrete charge carrier storage layer form the nonvolatile flash memory device (100).

WO 2006/112793 A1

NONVOLATILE FLASH MEMORY DEVICE AND METHOD FOR PRODUCING THE SAME

Field of the Invention

This invention relates generally to nonvolatile flash memory devices. In particular, this invention relates to a nonvolatile flash memory device comprising a discrete charge storage layer having metallic nanocrystals embedded in high-k dielectric material. The invention also relates to the fabrication of nonvolatile flash memory devices using metallic nanocrystals embedded in high-k dielectric material as a floating gate.

Background of the Invention

In recent years, the demand for high density and high performance nonvolatile memory devices has increased to meet the requirements of a range of modern microprocessor-based devices which currently run a variety of functions within cars, personal and portable computers, voice recorders, mobile phones, digital cameras, other wireless products, and industrial control systems. For the future, the "International Technology Roadmap of Semiconductors" (ITRS) 2003 shows that the scale down of nonvolatile flash memory devices will meet serious difficulty. In particular, scaling-down of the conventional polycrystalline silicon based flash memory cell size becomes more and more difficult, because the tunneling oxide thickness cannot be scaled down with the cell size due to the need to keep excellent charge retention and endurance. In order to solve this problem, new types of nonvolatile memory cells, such as discrete nanocrystal nodes (e.g. Si, Ge, SiGe and metal) based memory devices and trap based memory devices (e.g. SONOS, MONOS and SOHOS), are attracting a lot of research attention.

Discrete charge carrier storage resulting from nanocrystals (NCs) and Polysilicon-Oxide-Nitride-Oxide-Silicon (SONOS) type memory devices has been developed because the discreteness of charge carrier storage suppresses lateral migration of charge carriers, hence stored charge carriers are less vulnerable to oxide defects compared with conventional continuous floating gate memory devices.

Metal nanocrystals (NCs) are advantageous for memory devices based on NCs because of the wide range of available work functions, stronger coupling with the conduction channel and smaller energy perturbation due to large barriers for trapped charge carriers. According to simulation results using Wentzel-Kramer-Brillouin (WKB) approximation, leakage current can be reduced as much as two to four orders of magnitude using large work function metals as NCs. Conventionally, metals are forbidden in the front-end of line fabrication of nonvolatile

memory devices, but metal in the form of a NC is much more thermodynamically stable and can be integrated with less possibility to damage or contaminate the nonvolatile memory device.

For example, researchers at the National Science Foundation's Center for Advanced Materials and Smart Structures (CAMSS) demonstrated in 2004 arrays of magnetic nickel nanodots having a maximum dimension of 7 nm, and Motorola demonstrated also in 2004 a memory device based on silicon nanocrystals having a storage capacity of 4 MB.

Examples of nonvolatile memory devices of the NC- and SONOS-type are disclosed in the U.S. patents US 6,351,411 B2, US 6,407,424 B2, US 6,413,819 B1, US 6,545,314 B2, and US 6,724,038 B2.

In WO 2004/048923 A2 to Liu et al., metal NCs using platinum (Pt), silver (Ag), and gold (Au) on a gate oxide comprising silicon dioxide (SiO_2) have been investigated. These metals are chemically inert materials, but can be contaminants in silicon (Si) based fabrication during high temperature treatment. It can clearly be seen that the use of a metal already used in silicon technology is preferable. Since the refractory metal tungsten (W) is widely used as plug for contact metallization in integrated circuit (IC) fabrication, tungsten becomes a good candidate as material for NCs. The work function of tungsten (about 4.6 eV) is larger than the electron affinity of silicon or germanium (about 4.1 eV), resulting in a deeper quantum well. Hence, electrons captured by tungsten nanocrystals (W-NCs) do only escape with difficulty. The mid-gap work function of tungsten can be integrated for both N-MOSFETs and P-MOSFETs to control the threshold voltage. The high melting point and good conductivity of tungsten and the good compatibility of tungsten with silicon processing make W-NCs promising candidates among metal NCs for flash memory applications.

W-NCs embedded in silicon dioxide (SiO_2) based flash memory devices where the floating gate was fabricated by sputtering using a special target composition of dot and matrix materials have already been reported. However, thicker control and tunneling oxides are required to prevent floating gate charge carrier loss between the floating gate and the control gate or the floating gate and the substrate, thus resulting in the requirement of a large voltage to inject charge carriers into the floating gate. Therefore, programming efficiency is decreased due to the thick control and tunneling oxides and a high density of nanocrystals can degrade the device performance due to lateral migration of charge carriers, resulting in poor retention properties.

Direct tunneling through a thin tunneling oxide is most preferable for programming and erasing a nonvolatile flash memory device, but direct tunneling enhances the leakage current which reduces the retention time.

Currently, template technology using aluminum (III) oxide (Al_2O_3) or a polymer template is one of the competing nanocrystal fabrication processes to solve similar problems. However, the pore size of these templates is still large and their fabrication is not compatible with conventional CMOS process technology.

The object of this invention is to provide a nonvolatile flash memory device and a method for fabricating the same, which overcome the above mentioned shortcomings and which provide long retention time, high reliability, and compatibility with conventional CMOS process technology.

Summary of the Invention

To increase the programming efficiency and prevent charge loss, physically thick high-k dielectric materials with lower tunneling barriers and equivalent oxide thickness (EOT) are attractive alternatives to replace conventional silicon dioxide. The lower barrier height of high-k dielectric materials allows high currents across the tunneling oxide at low control gate voltages during programming and erasing the nonvolatile flash memory device. A thin tunneling oxide is also more vulnerable to metal contamination.

The invention provides a novel method for assembling metallic nanocrystals embedded in high-k dielectric material for a novel nonvolatile flash memory device using the metallic nanocrystals embedded in high-k dielectric material for storage of charge carriers, which largely improves the data retention and reliability of the nonvolatile flash memory device. The novel method of this invention can be performed with conventional CMOS process techniques. Therefore, the challenge presented in ITRS 2003 can be overcome.

In a first aspect the invention provides a nonvolatile flash memory device that comprises a discrete charge carrier storage layer. This discrete charge carrier storage layer comprises metallic nanocrystals as discrete storage nodes for discretely storing charge carriers, which metallic nanocrystals are embedded in high-k dielectric material. These metallic nanocrystals embedded in high-k dielectric material may be formed with the method according to the second aspect of this invention which is described below.

In the present invention, the term "high-k dielectric material" refers to a dielectric material having a dielectric constant k being clearly higher than 3.9 which is the value of the dielectric constant of silicon dioxide (SiO_2). In some embodiments of the invention, the high–k dielectric material has a dielectric constant k higher than about 5, about 6, about 8 or about 12. Therefore, the nonvolatile memory device of this invention employs metallic nanocrystals with high density and a suitable size distribution as discrete storage nodes for charge carriers. These metallic nanocrystals are insulated by high-quality high-k dielectric material and can be

formed using conventional CMOS process technology. Further, the present nonvolatile flash memory device has excellent retention properties in the discrete storage nodes represented by the metallic nanocrystals, while the barrier height from the metallic nanocrystals to the high-k dielectric material is lower than in prior art memory devices, and while the high-k dielectric material has a low EOT. This is *inter alia* due to the fact that charge carriers (electrons) are not stored in the metallic nanocrystals themselves but are stored in the quantum well caused by the metallic nanocrystals embedded in high-k dielectric material. Taken all this together enables to scale-down the programming voltage as well as the programming time compared with conventional memory devices.

In some embodiments, hafnium oxide-based or zirconium oxide-based high-k dielectric materials are used for embedding the metallic nanocrystals therein. A metal such as cobalt, gold, nickel, palladium, platinum, silver, tungsten or a metal containing compound such as titanium nitride is typically used for forming the metallic nanocrystals.

In one embodiment, hafnium aluminate (HfAlO) is considered one of the suitable high-k dielectric materials, because of its relatively high crystallization temperature ($\geq$ 1,000°C) and low leakage current. It has to be briefly noted here that hafnium aluminate (HfAlO) has a dielectric constant k of about 12 to about 18. Moreover, since HfAlO can be deposited by atomic layer deposition (ALD) with excellent control of both layer stoichiometry and thickness, HfAlO deposited by ALD can be used as a control and tunneling oxide. Other materials which have comparable properties, i.e. in particular a dielectric constant in the range of about 5 to about 12 or higher can also be used in which the metal containing nanocrystals are embedded.

Compared with the prior art, the retention and reliability of nonvolatile memory devices can be improved further by this invention. For this reason, the nonvolatile memory devices of this invention can also be used in an extreme environment such as, but limited to, astrospace and radiation.

This invention in addition provides the feasibility of double-bit storage capability, resulting in low cost per stored bit for flash memory products in industry. Further, this invention provides a solution for the programming, retention and reliability problems of the future scaled-down semiconductor flash memory devices, such as beyond 65 nm technology node. This has been marked as "Manufacturable solutions are NOT known" in ITRS 2003.

In a second aspect the invention provides a method of producing metallic nanocrystals embedded in high-k dielectric material. This method comprises the following steps: Firstly, a substrate covered with a first dielectric layer of a first high-k dielectric material is provided. Secondly, a second dielectric layer of a second high-k dielectric material is deposited over the

first dielectric layer. Then, a thin metal film, usually an ultra-thin metal film, is deposited over the second dielectric layer. Thereupon, this thin metal film is annealed. Metallic nanocrystals are formed, preferably in a self-assembled manner, from the (ultra-) thin metal film on the second dielectric layer during the annealing. Finally, the second dielectric layer and the metallic nanocrystals are covered with a third dielectric layer of the first high-k dielectric material. Thus, the metallic nanocrystals are embedded in high-k dielectric material.

The first, second and third dielectric layers together with the embedded metallic nanocrystals as the discrete charge carrier storage layer form the nonvolatile flash memory device according to the first aspect of the present invention.

It is noted that the method of the present invention can easily be incorporated in the common production of nonvolatile memory devices since this method is compatible to the conventional CMOS process technology, in particular to the front end processes thereof. This is especially due to the fact that no diffusion of the metallic nanocrystals is observed in the high-k dielectric material upon the annealing step. The annealing step is typically carried out at an elevated temperature, usually at an elevated below the melting point of the used materials.

Further embodiments are described and disclosed in the dependent claims and the detailed description, and will be also become apparent when reading the detailed description of this invention together with the accompanying drawings. The physical and electronic characteristics of one exemplary embodiment of one nonvolatile flash memory device according to the present invention are described in detail hereafter with respect to Fig.7 to Fig.14.

Brief Description of the Drawings

Fig.1 shows a schematic cross-sectional view of a nonvolatile flash memory device according to an embodiment of the invention.

Fig.2 illustrates a simplified fabrication process flow diagram for fabricating the nonvolatile flash memory device shown in Fig.1.

Fig.3A to Fig.3D show four SEM images of the uncompleted nonvolatile flash memory device shown in Fig.1 with different layers and/or in different fabrication states.

Fig.4A and Fig.4B show x-ray photoelectron spectra (XPS) depth profiles taken from layer stacks used in the nonvolatile flash memory device shown in Fig.1.

Fig.5A and Fig.5B show cross-sectional transmission electron microscopy (TEM) images of the nonvolatile flash memory device shown in Fig.1.

Fig.6A and Fig.6B show energy dispersive x-ray spectra (EDX) and a components list for two different spots indicated in Fig.5B.

Fig.7 shows the I-V transistor characteristics of the nonvolatile flash memory device shown in Fig.1.

Fig.8A and Fig.8B show the hysteresis loop of double sweep I-V and the gate to substrate C-V curves obtained from the nonvolatile flash memory device shown in Fig.1.

Fig.9A and Fig.9B show the transfer characteristics of the nonvolatile flash memory device shown in Fig.1 during programming/erasing and the threshold voltage shifts of the nonvolatile flash memory device shown in Fig.1 after applying programming/erasing voltage pulses.

Fig.10 shows the threshold voltage shifts of the nonvolatile flash memory device shown in Fig.1 as a function of programming voltage.

Fig.11 shows the endurance characteristics of the nonvolatile flash memory device shown in Fig.1 for program and erase operations.

Fig.12 shows the retention characteristics of the nonvolatile flash memory device shown in Fig.1 at room temperature.

Fig.13 shows a comparison table of the threshold voltage decay rates of nonvolatile flash memory devices of the present invention and of some prior art.

Fig.14 shows dual mode operations of the nonvolatile flash memory device shown in Fig.1 by charging source/drain regions and reading at the drain region in different programming/erasing conditions.

Detailed Description of the Invention

In the following detailed description of the invention, reference is made to the accompanying drawings which form a part hereof, and in which are shown, by way of illustration only, specific embodiments of this invention. In the drawings, like numerals describe substantially similar components throughout the several views.

The term substrate used in the following description refers to any doped and/or undoped semiconductor structure that has an exposed surface for the formation of an integrated circuit. Such a semiconductor structure may also comprise other layers that have been fabricated thereupon. Further, the terminals associated with the terms source and drain are determined by operating conditions of the nonvolatile flash memory device formed as a transistor, i.e. the terms source and drain are interchangeable. Additionally, the nonvolatile flash memory device described herein may be part of an arrayed memory, and may further comprise appropriate circuitry for driving and controlling the nonvolatile flash memory device. This circuitry as

such is well known in the art and therefore not described in detail herein. Furthermore, the term nanocrystals as used in the present invention means nanocrystalline particles each having a maximum dimension less than or equal to 10 nm, preferably between about 5 nm and about 7 nm, and in particular equal to 5 nm. These nanocrystalline particles are (physically) separated from each other, i.e. do not have contact with each other.

Referring now to **Fig.1**, a nonvolatile flash memory device 100 according to an embodiment of this invention is shown. The nonvolatile flash memory device 100 is formed on a substrate 101. A channel region 102 and source/drain regions 103, 104 on opposite sides of the channel region 102 are provided in the substrate 101. The channel region 102 and the source/drain regions 103, 104 are usually provided by suitable doping of appropriate regions of the substrate 101. Above the channel region 102, there are provided a first dielectric layer 105 and a second dielectric layer 106 covering the first dielectric layer 105. On the second dielectric layer 106 there are arranged metallic nanocrystals 107 in a self-assembled manner. These metallic nanocrystals 107 are distributed substantially two-dimensionally with substantially constant spaces on the second dielectric layer 106. Further, the metallic nanocrystals 107 and the regions of the second dielectric layer 106 left open by the metallic nanocrystals 107 are covered by a third dielectric layer 108. Finally, on top of the third dielectric layer 108 there is provided a control gate layer 109.

The first dielectric layer 105 and the second dielectric layer 106 together represent the tunneling oxide for the metallic nanocrystals 107, and the third dielectric layer 108 represents the control oxide for the metallic nanocrystals 107. Further, the metallic nanocrystals 107 constitute discrete charge carrier storage nodes in a discrete charge carrier storage layer of the nonvolatile flash memory device 100, wherein the discrete charge carrier storage nodes can be compared with a common floating gate layer sandwiched between gate oxide layers. The operation of the nonvolatile flash memory device 100 is known to a person of average skill in the art. Thus, no detailed description thereof is given herein.

According to the invention, the first, second and third dielectric layers 105, 106, 108 are formed of high-k dielectric material. In particular, the same high-k dielectric material may be used for the first and third dielectric layers 105, 108. The second dielectric layer 106 acts as a diffusion barrier layer for the metallic nanocrystals 107 for preventing diffusion of the metallic nanocrystals 107 within the first and third dielectric layers 105, 108. Further, the second dielectric layer 106 prevents contamination of the first dielectric layer 105 with metal of the metallic nanocrystals 107. Furthermore, the second dielectric layer 106 provides low elastic energy and high surface mobility such that the density of the metallic nanocrystals 107 can be increased compared with a device without the second dielectric layer 106.

According to one embodiment of this invention, the substrate 101 comprises silicon (Si), the source/drain regions 103, 104 comprises n^+-doped silicon (Si) using arsenic (As) as dopant, the metallic nanocrystals 107 comprise tungsten (W), and the first and third dielectric layers 105, 108 comprise hafnium aluminate (HfAlO) having a dielectric constant k_1 of about 12 to 18. In this embodiment, the second dielectric layer 106 comprises aluminum (III) oxide (Al_2O_3) having a dielectric constant k_2 of about 8, and the control gate layer 109 comprises tantalum nitride (TaN). In this embodiment the following alternative materials may also be used: for the substrate 101 strained silicon or silicon on insulator (SOI), for the source/drain regions 103, 104 as dopants phosphorus (P), and for the control gate layer 109 poly-silicon. Since tungsten (W) is used as material for the metallic nanocrystals 107 in the present embodiment, the term "W-NCs" is used as abbreviation for the metallic nanocrystals 107 in the following. However, the term "W-NCs" is by no means restricted to tungsten as the material used for the metallic nanocrystals 107, but other materials examples of which are presented in the following paragraph can also be employed. The metallic nanocrystals 107 typically have a maximum dimension, i.e. an average size, less than or equal to 10 nm, preferably between about 5 nm and about 7 nm, and in particular equal to 5 nm. The metallic nanocrystals 107 are distributed in a self-assembled manner on the second dielectric layer 106 due to the method for forming the metallic nanocrystals 107 as described below with respect to Fig.2. Further, the metallic nanocrystals 107 are distributed as floating gate in the discrete charge carrier storage layer with a density of about $5\times10^{11}/cm^2$. The average size of the metallic nanocrystals 107 and their density are close to the demand of ITRS 2003.

In accordance with the above explanations other materials usable in other embodiments of this invention include, but are not limited to: For the first and third dielectric layers 105, 108 any hafnium oxide-based or zirconium oxide-based high-k dielectric material can for examples be used since hafnium and zirconium belong to the same main group in the periodic table of the elements. Examples of suitable materials that can be used instead of hafnium aluminate (HfAlO) include, but are not limited to, hafnium silicate (HfSiO) having a dielectric constant k of about 3.9 to 26, hafnium silicon oxynitride (HfSiON) having a dielectric constant k of about 5 to 25, or zirconium oxide (ZrO_2) having a dielectric constant k in the range of 20 to 25. Examples of suitable materials that can be used as material for the second dielectric layer 106 instead of aluminum (III) oxide (Al_2O_3) include, but are not limited to, a high-k dielectric material such as silicon (IV) nitride (Si_3N_4) having a dielectric constant k in the range of about 6 to 7, or titanium (IV) oxide (TiO_2) having a dielectric constant k in the range of 20 to 85. Finally, examples metal containing material or metals other than tungsten that can

be used for the formation of the metallic nanocrystals 107 include, but are not limited, to cobalt, gold, nickel, palladium, platinum, silver or titanium nitride.

Fig.2 illustrates a simplified fabrication process flow diagram for fabricating the nonvolatile flash memory device 100 shown in Fig.1.

In step S201, a bare silicon substrate 101 is provided, which was cleaned by a standard process first. Then, the first dielectric layer 105 is deposited with a thickness in the range of 3 nm to 7 nm, preferably with a thickness of about 5 nm, on the substrate 101 in step S202 by atomic layer deposition (ALD). Since the first dielectric layer 105 comprises hafnium aluminate (HfAlO) in this embodiment of the invention, the used ALD process is based on alternating pulses of the precursors hafnium tetrachloride ($HfCl_4$) and water (H_2O) on the one hand, and tri-methyl aluminum ($Al(CH_3)_3$) and water (H_2O) on the other hand. In step S203, the second dielectric layer 106 is deposited *in-situ* with a thickness in the range of 0.5 nm to 1.5 nm, preferably with a thickness of about 1 nm, on the first dielectric layer 105 also by ALD without breaking vacuum. Since the second dielectric layer 106 comprises aluminum (III) oxide (Al_2O_3) in this embodiment of the invention, the used ALD process is based on pulses of the precursors tri-methyl aluminum ($Al(CH_3)_3$) and water (H_2O) only. Since aluminum (III) oxide (Al_2O_3) has a higher conduction band offset (experimentally 2.37±0.15 eV and theoretically 2.8 eV) compared to the conduction band offset of hafnium aluminate (HfAlO) (2.1 eV) such that low voltage operation of the nonvolatile flash memory device 100 would not be possible, it has to be guaranteed that the thickness of the second dielectric layer 106 is sufficiently low such that the above mentioned effects do not take effect. The inventors of this invention found that a thickness of about 1 nm is sufficient to allow low voltage operation of the nonvolatile flash memory device 100 as well as to prevent diffusion of the metallic nanocrystals 107 within the first and third dielectric layers 105, 108.

A thin metal layer, typically an ultra-thin metal layer, also named wetting layer, substantially consisting of the metal tungsten (W) is then sputtered with a thickness of between 1 nm and 5 nm onto the second dielectric layer 106 in step S204. A sputtering period of between 10 sec and 60 sec, in particular of about 30 sec, a power of 50 W and a pressure of 0.4 Pa (about 3 mT) were used in this sputtering process. As an alternative to sputtering, electron beam evaporation (e-beam evaporation) and atomic layer chemical vapor deposition (ALCVD) can be used for depositing the ultra-thin metal film onto the second dielectric layer 106.

In this fabrication process of the nonvolatile flash memory device 100, the step S204 is immediately followed by step S205, wherein the ultra-thin metal layer is annealed by rapid thermal annealing (RTA). During this RTA in step S205, metallic nanocrystals 107 are formed

in a self-assembled manner from the ultra-thin metal layer. Since self-assembly is a thermodynamic process, the density, size and distribution of the metallic nanocrystals 107 can be controlled by optimizing the initial thickness of the ultra-thin metal film, the annealing temperature used for RTA, and the annealing time during which RTA takes place. According to one embodiment of this invention, the RTA is typically carried out in an inert ambient. Any ambient that is chemically inert (unreactive) towards the used materials can be used. The atmosphere can entirely consist or substantially comprise an inert gas such as a noble gas (e.g. xenon, argon) or nitrogen (N_2) or a mixture of such inert gases. Depending on the used materials, the RTA is usually carried out at a temperature between about 500°C and about 1,000°C, for example between about 900°C and about 1,000°C, and in one embodiment in which the dielectric layer comprises HfAlO at 950°C. In this last mentioned and other embodiments, the RTA is typically carried out for a time of about between 20 sec and about 2 min, in particular for 60 sec. However, the RTA is by no means limited to the above conditions and can be carried out under any suitable reaction conditions which lead to the formation of nanocrystals that are suitable for practicing the present invention. The determination of suitable annealing conditions is within the knowledge of the person of average skill in the art and can be carried out empirically. This technique as used in the present invention can also be applied to combinations of metallic nanocrystals and high-k substrates, which include but are not limited to, e.g., Au/HfO_2, Au/HfAlO, Au/Al_2O_3, Ag/HfO_2, Ag/HfAlO, Ag/Al_2O_3, Pt/HfO_2, Pt/HfAlO, Pt/Al_2O_3. In addition, the uniformity and size distribution of the metallic nanocrystals 107 can be improved by using suitable equipment such that a highly uniform ultra-thin metal layer is provided. Since the first dielectric layer 105 has and maintains an amorphous structure regardless of any temperature treatment, the surface energy of the metallic nanocrystals 107 on the second dielectric layer 106 is minimized. As already mentioned above with respect to Fig.1, the metallic nanocrystals 107 are distributed on the second dielectric layer 106 with a density of about $5\times10^{11}/cm^2$. If one should refrain from using the second dielectric layer 106 and should form the metallic nanocrystals 107 directly on the first dielectric layer 105, only a density of about $1\times10^{11}/cm^2$ can be achieved for the metallic nanocrystals 107. It should be noted that if the ultra-thin metal layer has a thickness greater than 5 nm, then the self-assembly of metallic nanocrystals 107 may not take place due to a too high lattice binding energy of the metal in the ultra-thin metal layer.

Following the step S205, a third dielectric layer 108 is deposited with a thickness of between 8 nm and 10 nm over the metallic nanocrystals 107 and the second dielectric layer 106 in step S206 by ALD. The step S206 is comparable with step S202. Since the third dielectric layer 108, like the first dielectric layer 105, comprises hafnium aluminate (HfAlO)

in this embodiment of the invention, the used ALD process is also based on alternating pulses of the precursors hafnium tetrachloride ($HfCl_4$) and water (H_2O) on the one hand, and tri-methyl aluminum ($Al(CH_3)_3$) and water (H_2O) on the other hand.

The step S206 is followed by the step S207, wherein a control gate layer 109 comprising tantalum nitride (TaN) is deposited with a thickness of about 150 nm over the third dielectric layer 108 by sputtering. Tantalum nitride (TaN) is used as material for the control gate layer 109 due to its good thermal stability and electrical characteristics when used along with the high-k dielectric materials used in this embodiment of this invention.

The following step S208 represents an appropriate process for lithographically exposing the layer stack comprising the discrete charge carrier storage layer (i.e., having the first, second and third dielectric layers 105, 106, 108 and the metallic nanocrystals 107) and the control gate layer 109. In step S209, this lithographically exposed layer stack is then etched in a suitable etch process for laterally confining the layer stack comprising the trapping layer, and for forming uncovered substrate regions on opposite sides of the laterally confined layer stack. Afterwards, arsenic (As) is implanted into the uncovered substrate regions in step S210 for forming source and drain regions 103, 104 by an implantation process using arsenic ions at an energy of about 100 keV and an area dose of about $1\times10^{15}/cm^2$. The source and drain regions 103, 104 are finally activated in step S211 by RTA at a temperature of about 950°C and a duration of about 30 sec. Therefore, the source and drain regions 103, 104 consist of n^+-doped silicon, wherein the n^+-dopant is arsenic. Then, the nonvolatile flash memory device 100 is completed.

For completeness, it is mentioned here that the resulting transistor structure, which also forms the basis for all following test results, has a width of about 100 μm and a channel length of about 20 μm.

It has to be noted that each individual process used in the above-described steps is commonly known in CMOS production technology. Therefore, the fabrication of the nonvolatile flash memory device 100 of this invention can be easily integrated into common CMOS production.

As every person skilled in the art knows, a plurality of nonvolatile flash memory devices 100 can also be arranged in a memory device array further comprising suitable electronics for storing into and reading from the individual nonvolatile flash memory devices 100 in such a memory device array.

Fig.3A to Fig.3D show four scanning electron microcopy (SEM) images of the uncompleted nonvolatile flash memory device 100 (compare Fig.1) with different layers and/or in different fabrication states.

Fig.3A shows an SEM image of the uncompleted nonvolatile flash memory device 100 after step S204, i.e. before RTA. In particular, Fig.3A shows solely the extended surface of the ultra-thin metal film. Further, it is shown in Fig.3A that metallic nanocrystals 107 cannot be formed without annealing.

Fig.3B and Fig.3C show SEM images of the uncompleted nonvolatile flash memory device 100 after step S205, i.e. after RTA. The difference between Fig.3B and Fig.3C is that for Fig.3B step S203 has been skipped whereas for Fig.3C step S202 has been skipped. I.e., Fig.3B shows metallic nanocrystals 107 directly on the first dielectric layer 105 without any second dielectric layer 106 in between, and Fig.3C shows metallic nanocrystals 107 on the second dielectric layer 106 deposited directly on the substrate 101 without any first dielectric layer 105. As can be gathered from Fig.3B and Fig.3C, more uniform metallic nanocrystals 107 with a higher density ($5\times10^{11}/cm^2$) can be observed on the second dielectric layer 106 (comprising aluminum (III) oxide (Al_2O_3) than on the first dielectric layer 105 (comprising hafnium aluminate (HfAlO) and a density of about $1\times10^{11}/cm^2$). However, as already mentioned above, low voltage operation would not be possible if only aluminum (III) oxide (Al_2O_3) is used as high-k dielectric material, since the conduction band offset of aluminum (III) oxide (Al_2O_3) would be too high (2.8 eV) compared to the conduction band offset (2.1 eV) of hafnium aluminate (HfAlO). Therefore, the present invention combines these two effects, i.e. provides a low voltage operation and a high density of the metallic nanocrystals 107 by providing the first and second dielectric layers 105, 106 as described above for Fig.1.

Fig.3D shows a SEM image of the uncompleted nonvolatile flash memory device 100 after step S205, i.e. after RTA, with all other previous steps executed. I.e., Fig.3D shows metallic nanocrystals 107 on a layer stack comprising directly below the metallic nanocrystals 107 the second dielectric layer 106 and underneath the first dielectric layer 105. As can be gathered from Fig.3D, the metallic nanocrystals 107 distributed on a $Si/HfAlO/Al_2O_3$ layer stack have a high density which is comparable to metallic nanocrystals 107 distributed solely on a Si/Al_2O_3 layer stack.

Therefore, the nonvolatile flash memory device 100 of this invention provides good charge carrier storage capabilities with good uniformity as well as low programming efficiency, since the metallic nanocrystals 107 are not too small or dispersed. Further, the nonvolatile flash memory device 100 of this invention provides good charge carrier insulation capabilities, since movement of charge carriers due to too large or dense metallic nanocrystals 107, or leakage of charge carriers due to defects in the tunneling or control oxide are avoided. In the nonvolatile flash memory device 100 of this invention, a low programming voltage is achieved using a high-k dielectric material with small conduction band offset, but thin EOT,

whereas the charge retention characteristics are improved by using a physically thick high-k dielectric material based on an oxide.

According to this invention, feasibility of double-bit storage capability is shown by selectively injecting charge carriers by means of channel hot electron injection (CHEI) at the source region, at the drain region, and at both source and drain regions.

Fig.4A and Fig.4B show x-ray photoelectron spectra (XPS) depth profiles taken from layer stacks used in the nonvolatile flash memory device 100 shown in Fig.1. In particular, **Fig.4A** shows the XPS depth profile for a layer stack of Si/HfAlO/W-NCs/HfAlO, whereas **Fig.4B** shows that for a layer stack of Si/HfAlO/Al_2O_3/W-NCs/HfAlO (ls1 and ls2 are chosen for a linear square fit during the data analysis process). It can be gatherer from Fig.4A that there is no distinct second dielectric layer 106 (Al_2O_3) between the W-NCs 107 and the first dielectric layer 105 (HfAlO) after the high temperature annealing for the formation of the self-assembled W-NCs 107. Further, from Fig.4B it can be gathered that there exists an Al rich layer between the first dielectric layer 105 (HfAlO) and the W-NCs 107. The Al rich layer serves to increase the uniformity and density of the metallic nanocrystals 107, and also lowers the programming voltage compared with Al_2O_3 due to the lower conduction band offset of HfAlO.

Fig.5A and Fig.5B show cross-sectional transmission electron microscopy (TEM) images of the nonvolatile flash memory device 100 shown in Fig.1. **Fig.5A** shows the TEM image of a Si/HfAlO/W-NCs/HfAlO layer stack where spherical W-NCs 107 are formed according to the fabrication process described for Fig.2 in this invention. A thin interfacial layer (IL) 501 comprising hafnium silicate (HfSiO) between the first dielectric layer 105 and the substrate 101 is formed due to the high temperature treatment during device processing. The TEM image shows clearly the amorphous structure of the first and third dielectric layers 105, 108 (HfAlO) for both the tunneling and control oxides after high temperature treatment at 950°C for 60 sec. **Fig.5B** shows an enlarged high-resolution TEM image of Fig.5A, in which the lattice structure of the W-NCs 107 extending over about 5 nm can clearly be seen. The inset in Fig.5A shows the regular pattern of a single electron diffraction (SED) image of one W-NC, which shows that the metallic nanocrystals 107 are really crystalline. Therefore, high quality metallic nanocrystals 107 in terms of shape and crystallinity can be produced by the self-assembly process described in this invention.

Fig.6A and Fig.6B show energy dispersive x-ray spectra (EDX) and a components list for two different spots indicated in Fig.5B. **Fig.6A** shows the possible diffusion of the metal tungsten (W) of the metallic nanocrystals 107 into the control and tunneling oxides using EDX analysis. A strong tungsten (W) signal was detected at the metallic nanocrystals 107 (spot 1),

whereas no tungsten (W) signal was found at the interfacial dielectric layers 105, 501 (spot 2). Also, no diffusion of tungsten (W) is observed into the second dielectric layer 106 (Al_2O_3) and into the Al rich first dielectric layer 105 (HfAlO) (not shown).

Fig.6B shows a components list for the two different spots indicated in Fig.5B. It can clearly be gathered from this elements list that tungsten is present in spot 1 with a portion of 34.60 weight-% and 12.50 atomic-%, whereas no tungsten (W) is present at spot 2.

Prior studies have reported that most of platinum (Pt) diffused through an underlying 3 nm thick silicon dioxide (SiO_2) interface during the formation of platinum nanocrystals by annealing at 850-900°C forms platinum silicide due to reacting with the silicon (Si) of the substrate. According to the EDX analysis in this application, it appears that the diffusion of tungsten (W) is self-limited during the formation of the nanocrystals. Furthermore, it is expected that diffusion of higher melting point metals should be negligible during annealing for the formation of nanocrystals. The EDX result clearly shows that W-NCs are thermodynamically stable, non-diffusive into high-k dielectric materials, and compatible with front end processes.

The well behaved I_{DS}-V_{DS} transistor characteristics of a nonvolatile flash memory device 100 (shown in Fig.1) based on an N-MOS transistor with W-NCs embedded in hafnium aluminate (HfAlO) are shown in **Fig.7**, the I_{DS}-V_{DS} transistor characteristics being shown for V_{GS} = 1 V, 2 V and 3 V. Therein, V_{DS} and I_{DS} stand for the voltage and the current, respectively, between the drain and source regions 103, 104, and V_{GS} stands for the voltage between the control gate layer 109 and the source region. It can easily be gathered from Fig.7 that I_{DS} increases with increasing V_{GS}, and that I_{DS} saturates with increasing V_{DS}. **Fig.8A** shows the logarithmic I_D-V_G characteristics of the nonvolatile flash memory device 100 shown in Fig.1 with gate voltage double sweep ranges of ±2 V, ±3 V, ±4 V and ±5 V. Therein, V_G stands for the gate voltage, and I_D and V_D stand for the current through and the voltage at the drain region, respectively. The threshold voltage shift between the forward and reverse sweeps was found to be 2 V for ±5 V. The shift of I_D-V_G curve during the reverse sweep is larger compared with the forward one and the threshold voltage saturation is observed at beyond 3 V during the forward sweep. This indicates that the programming is dominated by electron trapping into W-NCs for each case. The same phenomenon is observed in the high frequency gate to substrate C-V hysteresis curves which are shown in **Fig.8B** for the nonvolatile flash memory device 100 shown in Fig.1. Therein, C/C_{ox} stands for the normalized capacitance of the nonvolatile flash memory device 100 shown in Fig.1.

Fig.9A shows the I_D-V_G transfer characteristics of the nonvolatile flash memory device 100 shown in Fig.1 during programming (+)/erasing (-) using voltage pulses of ±7 V with up

to 1 sec pulse width. Further, **Fig.9B** shows the threshold voltage shifts ΔV_{th} of the nonvolatile flash memory device 100 shown in Fig.1 after applying programming (-5 V and -7 V)/erasing (+5 V and +7 V) voltage pulses.

Symmetrical programming and erasing characteristics are observed from Fig.9A and Fig.9B. The programming/erasing window is found to be 1.8 V for ±7 V programming/erasing voltage pulses. If the programming speed is defined as the time to reach 90% of ΔV_{th}, the programming speed in this work is 1 μs at 7 V operating bias which is close to the ITRS 2003 requirement for beyond 2006, whereas in case of W-NCs in silicon dioxide (SiO_2) with a physical thickness of the tunneling oxide of 5 nm the programming speed is about 100 ms at 10 V operating bias. This indicates that the programming speed of W-NCs in hafnium aluminate (HfAlO) is faster than of W-NCs in silicon dioxide (SiO_2). The conduction band offset of the hafnium aluminate (HfAlO) tunneling oxide is 2.1 eV which is lower than the conduction band offset of silicon dioxide (SiO_2) being 3.1 eV. This low electron barrier height enhances the electron tunneling through a triangular potential barrier by Fowler-Nordheim (F-N) conduction mechanism at low programming voltage and a larger part of the applied control gate voltage drops across the tunneling oxide due to the low equivalent oxide thickness (EOT) of hafnium aluminate (HfAlO). This enables low voltage and fast programming in this invention. The threshold voltage shift saturates as the programming/erasing time and voltage increases. This can be due to the presence of a finite number of W-NC nodes embedded in hafnium aluminate (HfAlO) and each node accepts only a finite number of charge carriers (here: electrons) under this testing condition. From the threshold voltage shift after programming at 5 V for 1 μs, the density of charge carriers ($Q_t = (\varepsilon / t_{control}) \Delta V_{th}$, where ε is the dielectric constant of hafnium aluminate (HfAlO) and $t_{control}$ is the thickness of the control oxide) is found to be $3.3\times10^{12}/cm^2$. Considering a NC density of $5\times10^{11}/cm^2$ estimated from the SEM image shown in Fig.3D, about seven electrons can be injected into a single W-NC after programming at 5 V for 1 μs. According to the simulation results reported in the prior art, about ten electrons can be confined in a silicon quantum dot (size: 10 nm×10 nm×6 nm) after 4 V programming for 1 μs. Therefore, the experimental result of this invention is comparable with the simulated result in the prior art.

It is observed from Fig.9B that the erasing speed is lower than the programming speed under both bias conditions. At thermal equilibrium, the band is aligned such that the mid-gap work function of the metal tungsten (W) creates a deep quantum well (2.6 eV) compared to the conduction band offset of the silicon (Si) substrate (2.1 eV) and the conduction band edge is higher than that of the W-NC. Charge carriers (here: electrons) can be easily injected into the quantum well during programming. On the contrary, the same amount of threshold voltage

change during erasing is possible by applying a higher erasing voltage than the programming voltage for the same reason. But in the erasing operation, the abrupt change in threshold voltage is observed after erasing the nonvolatile flash memory device for a long period. This abrupt change takes place earlier under the higher voltage erasing condition. This is not fully understood yet. One possibility may be the abrupt increase in hole capture efficiency in the quantum well with increasing erasing voltage.

In **Fig.10**, the programming characteristics (i.e. the threshold voltage shifts) of the nonvolatile flash memory device 100 shown in Fig.1 as a function of programming voltage are compared with a nonvolatile flash memory device without W-NCs. The devices were programmed at a write voltage V_G of between 1 V and 7 V for 1 ms. For the nonvolatile flash memory device 100 shown in Fig.1, the threshold voltage shift ΔV_{th} increases with increasing programming voltage V_G. A small threshold voltage shift ΔV_{th} was also observed at a high voltage in the nonvolatile flash memory device without W-NCs due to traps in the high-k dielectric material. It is found that the nonvolatile flash memory device of this invention (i.e. with W-NCs) gives a threshold voltage shift ΔV_{th} which is 30 times larger than for the nonvolatile flash memory device without W-NCs. This clearly shows that W-NCs are playing a major role during programming.

Fig.11 shows the endurance characteristics of the nonvolatile flash memory device 100 shown in Fig.1 for program (write or W) and erase (E) operations to assess the wear-out performance of the nonvolatile flash memory device 100. The endurance was tested using voltage pulses of ±7 V and pulse width of 10 ms per cycle for programming and erasing. In particular, Fig.11 shows the program/erase threshold voltage window as a function of the number of program/erase cycles. The window remains wide, although a small shift is observed after 10^5 cycles, perhaps due to some charge trapping/detrapping in the high-k dielectric material.

The retention characteristics of the nonvolatile flash memory device 100 shown in Fig.1 at room temperature is illustrated in **Fig.12**. The threshold voltage decay rate is found to be 50 mV/dec after programming with a pulse voltage of 7 V for 1 ms pulse width. The extrapolated memory window is still open at 0.22 V for a retention time of 10 years which is comparable or better with any known discrete nanocrystal memory device. The charge loss observed until 10^4 sec is much better than the results of W-NCs embedded in tunneling oxide comprising silicon dioxide (SiO_2).

Fig.13 shows a comparison table of the threshold voltage decay rates of nonvolatile flash memory devices of this invention and of some prior art (WO 2004/048923 A2; Takata M., Kondoh S., Sakaguchi T., Choi H., Shim J.C., Kurino H., and Koyanagi M.: "New non-

volatile memory with extremely high density metal nano-dots" in *Electron Devices Meeting,* 2003, pp. 22.5.1 - 22.5.4; Lee C., Gorur-Seetharam A., and Kan E.C.: "Operational and reliability comparison of discrete-storage nonvolatile memories: advantages of single- and double-layer metal nanocrystals" in *Electron Devices Meeting,* 2003, pp. 22.6.1 - 22.6.4). In particular, Fig.13 shows that the retention properties of the nonvolatile flash memory device 100 of this invention are better than or at least comparable with that of the different prior art memory devices having metal NCs.

Charge loss from NCs during retention can originate from a stress induced leakage current (SILC) and/or a direct tunneling current through the tunneling oxide, and from a current laterally tunneling from NC to NC. According to prior art simulation results, the potential of a dot is still near the mid-gap of a silicon (Si) channel during the retention mode, and the leakage process of stored electrons requires thermal excitation. Lateral charge distribution due to hopping of charge carriers between adjacent NCs may be one of the causes of threshold voltage decay. A larger memory window can be expected by optimizing the process condition for decreasing the traps in the high-k dielectric material and by designing the device structure for decreasing the leakage current through both the control and the tunneling oxides.

Fig.14 shows dual mode operations of the nonvolatile flash memory device 100 shown in Fig.1 by charging source/drain regions and reading at the drain region in different programming/erasing conditions.

Compared with a conventional single-bit cell, a dual-bit cell increases the cell density in a memory device array by reducing the memory cell area from two virtual transistors in place of one. The feasibility of a dual-bit cell in a memory device can be shown through channel hot electron injection (CHEI) programming / Fowler-Nordheim (F-N) erasing with forward/reverse reading, or through local charge storing by CHEI at the source region, at the drain region, and at both the source and drain regions. Since the source and drain regions 103, 104 are symmetrical in the structure of the nonvolatile flash memory device 100 shown in Fig.1, one of n^+-doped regions where positive voltage is applied during threshold voltage measurement is considered as drain region in the feasibility test of a two bit storage experiment. In the nonvolatile flash memory device 100 of this invention, charge carriers are locally stored through CHEI at the drain region (V_D = 5 V, V_G = 7 V), at the source region (V_S = 5 V, V_G = 7 V), and at both the source and drain regions (V_S = 5 V, V_G = 7 V followed by V_D = 5 V, V_G = 7 V). During CHEI programming, a lateral electric field near the source/drain regions 103, 104 heats the electrons and a transverse electric field between the channel region 102 and the control gate layer 109 injects the electrons into the W-NCs 107.

Erasing operation is performed by two-sided hot hole injection (HHI) resulting from band to band tunneling with positive voltages at the source/drain regions 103, 104 and a negative voltage at the control gate layer 109 (V_S = 5 V, V_G = -7 V, V_D = 5 V). As a result, holes are injected through the tunneling oxide into the W-NCs, where the stored electrons are neutralized by hot holes. The V_{th} shown in Fig.14 was measured by applying a positive bias (1 V) at the drain region and stepping the gate voltage. When electrons are stored at the source region and reading is taken at the drain region, charge carriers localized at the source region will screen the channel region near the source region due to a strong lateral electric field generated by the drain bias and as a result, V_{th} (S) is greater than V_{th} (D). Due to the storage of charge carriers at both the source and the drain regions, the threshold voltage is at the largest when compared with other charge storage conditions and at the lowest with no charge carriers stored in the trapping layer after HHI. Therefore, four well distinct states are present, which correspond to no charge carriers stored, charge carriers stored only on one side, and charge carriers stored on both sides. The threshold voltage separation of different states can be enhanced by controlling the channel doping. From the results discussed here it is concluded that the high density W-NCs in the trapping layer are well isolated by high-k dielectric material. Due to the high work function of the W-NCs, they are immune to lateral charge migration upon thermal emission from 1 eV to 2 eV shallow traps in a nitride storage layer causing charge migration, which is a clear advantage over known SONOS- or NROM-devices with respect to device scaling. Therefore, it is considered that nonvolatile flash memory devices of this invention with metallic nanocrystals embedded in high-k dielectric material can be used as a two bit-per-cell storage device in a single transistor.

In another embodiment of this invention, a double layer of metallic nanocrystals may be provided above the first and second dielectric layers. This enhances more retention and multi-bit operation properties for such a nonvolatile flash memory device of this invention.

Although this invention has been described in terms of preferred embodiments, it will be understood that numerous variations and modifications may be made, without departing from the spirit and scope of this invention as set out in the following claims.

Claims

What is claimed is:

1. A nonvolatile flash memory device comprising a discrete charge carrier storage layer, the discrete charge carrier storage layer comprising metallic nanocrystals embedded in high-k dielectric material.

2. The nonvolatile flash memory device as claimed in claim 1, wherein the discrete charge carrier storage layer comprises a layer stack successively having a first dielectric layer, a second dielectric layer and a third dielectric layer, each of the first, second and third dielectric layers comprising high-k dielectric material, and wherein the metallic nanocrystals are embedded between the second dielectric layer and the third dielectric layer.

3. The nonvolatile flash memory device as claimed in claim 2, wherein the first and third dielectric layers comprise a hafnium oxide-based or zirconium oxide-based high-k dielectric material.

4. The nonvolatile flash memory device as claimed in claim 2 or 3, wherein the first dielectric layer comprises a thickness between about 3 nm and about 7 nm.

5. The nonvolatile flash memory device as claimed in any of claims 2 to 4, wherein the first dielectric layer comprises a thickness of about 5 nm.

6. The nonvolatile flash memory device as claimed in any of claims 2 to 5, wherein the second dielectric layer comprises a thickness between about 0.5 nm and about 1.5 nm.

7. The nonvolatile flash memory device as claimed in claim 6, wherein the second dielectric layer comprises a thickness of about 1 nm.

8. The nonvolatile flash memory device as claimed in any of claim 2 to 7, wherein the third dielectric layer comprises a thickness between about 8 nm and about 10 nm.

9. The nonvolatile flash memory device as claimed in any of claims 1 to 8, wherein the discrete charge carrier storage layer is arranged above a substrate.

10. The nonvolatile flash memory device as claimed in claim 9, wherein the substrate comprises silicon.

11. The nonvolatile flash memory device as claimed in claim 9 or 10, wherein a control gate layer is located above the discrete charge carrier storage layer, wherein a channel region is located between the discrete charge carrier storage layer and the substrate, and wherein source and drain regions are located on opposite sides of the channel region on and/or in the substrate.

12. The nonvolatile flash memory device as claimed in any of claims 2 to 11, wherein the second dielectric layer is a diffusion barrier layer for the metallic nanocrystals.

13. The nonvolatile flash memory device as claimed in any of claims 2 to 12, wherein the first and third dielectric layers comprise a first high-k dielectric material selected from the group consisting of hafnium aluminate (HfAlO), hafnium silicate (HfSiO), hafnium silicon oxynitride (HfSiON), and zirconium oxide (ZrO_2).

14. The nonvolatile flash memory device as claimed in any of claim 2 to 13, wherein the second dielectric layer comprises a second high-k dielectric material selected from the group consisting of aluminum (III) oxide (Al_2O_3), silicon (IV) nitride (Si_3N_4), and titanium (IV) oxide (TiO_2).

15. The nonvolatile flash memory device as claimed in any of claims 1 to 14, wherein the metallic nanocrystals comprise a material selected from the group consisting of cobalt, gold, nickel, palladium, platinum, silver, tungsten, and titanium nitride.

16. The nonvolatile flash memory device as claimed in any of claims 1 to 15, wherein the metallic nanocrystals comprise a maximum dimension of less than or equal to 10 nm.

17. The nonvolatile flash memory device as claimed in claim 16, wherein the metallic nanocrystals comprise a maximum dimension of between about 5 nm and about 7 nm.

18. The nonvolatile flash memory device as claimed in claim 17, wherein the metallic nanocrystals comprise a maximum dimension of equal to 5 nm.

19. The nonvolatile flash memory device as claimed in any of claims 1 to 18, wherein the metallic nanocrystals are distributed in a self-assembled manner.

20. The nonvolatile flash memory device as claimed in any of claims 1 to 19, wherein the metallic nanocrystals are distributed substantially two-dimensionally in the discrete charge carrier storage layer.

21. The nonvolatile flash memory device as claimed in any of claims 1 to 20, wherein the metallic nanocrystals are distributed in the discrete charge carrier storage layer with a density of about $5\times10^{11}/cm^2$.

22. A method of producing metallic nanocrystals embedded in high-k dielectric material, comprising:

a) providing a substrate covered with a first dielectric layer of a first high-k dielectric material;

b) depositing a second dielectric layer of a second high-k dielectric material over the first dielectric layer;

c) depositing an ultra-thin metal film over the second dielectric layer;

d) annealing the ultra-thin metal film, thereby forming metallic nanocrystals on the second dielectric layer; and

e) covering the second dielectric layer and the metallic nanocrystals with a third dielectric layer of the first high-k dielectric material, thereby forming metallic nanocrystals embedded in high-k dielectric material.

23. The method as claimed in claim 22, wherein step d) is carried out in a substantially inert ambient.

24. The method as claimed in claim 22 or 23, wherein step d) is carried out in an ambient substantially comprising nitrogen.

25. The method as claimed in any of claims 22 to 24, wherein step d) is carried out at a temperature between about 500°C and about 1,000°C.

26. The method as claimed in any of claims 22 to 25, wherein step d) is carried out at a duration of between 20 sec and 2 min.

27. The method as claimed in any of claims 22 to 26, wherein the first high-k dielectric material used in steps a) and e) is a hafnium oxide-based or zirconium oxide-based high-k dielectric material.

28. The method as claimed in claim 27, wherein the first high-k dielectric material used in steps a) and e) is selected from the group consisting of hafnium aluminate (HfAlO), hafnium silicate (HfSiO), hafnium silicon oxynitride (HfSiON), and zirconium oxide (ZrO_2).

29. The method as claimed in any of claims 22 to 28, wherein the second high-k dielectric material used in step b) is selected from the group consisting of aluminum (III) oxide (Al_2O_3), silicon (IV) nitride (Si_3N_4), and titanium (IV) oxide (TiO_2).

30. The method as claimed in any of claims 22 to 29, wherein in step c) a metallic material is used, which metallic material is selected from the group consisting of cobalt, gold, nickel, palladium, platinum, silver, tungsten, and titanium nitride.

31. The method as claimed in any of claims 22 to 30, wherein in step d) metallic nanocrystals are formed, the maximum dimension of which is less than or equal to about 10 nm.

32. The method as claimed in claim 31, wherein in step d) metallic nanocrystals are formed, the maximum dimension of which is between about 5 nm and about 7 nm.

33. The method as claimed in claim 32, wherein in step d) metallic nanocrystals are formed, the maximum dimension of which is equal to 5 nm.

34. The method as claimed in any of claims 22 to 33, wherein in step d) metallic nanocrystals are formed, which are distributed in a self-assembled manner.

35. The method as claimed in any of claims 22 to 34, wherein in step d) metallic nanocrystals are formed, which are distributed substantially two-dimensionally on the second dielectric layer.

36. The method as claimed in any of claims 22 to 35, wherein in step d) metallic nanocrystals are formed, which are distributed on the second dielectric layer with a density of about $5\times10^{11}/cm^2$.

37. The method as claimed in any of claim 22 to 36, wherein the first dielectric layer in step a) is provided with a thickness between about 3 nm and about 7 nm.

38. The method as claimed in claim 37, wherein the first dielectric layer in step a) is provided with a thickness of about 5 nm.

39. The method as claimed in any of claims 22 to 38, wherein the first dielectric layer is provided in step a) by atomic layer deposition of the first high-k dielectric material onto the substrate.

40. The method as claimed in any of claims 22 to 39, wherein the second dielectric layer is deposited in step b) with a thickness between about 0.5 nm and about 1.5 nm.

41. The method as claimed in claim 40, wherein the second dielectric layer is deposited in step b) with a thickness of about 1 nm.

42. The method as claimed in any of claims 22 to 41, wherein the second dielectric layer is deposited in step b) by atomic layer deposition of the second high-k dielectric material onto the first dielectric layer.

43. The method as claimed in any of claims 22 to 42, wherein the ultra-thin metal layer is deposited in step c) with a thickness of between about 1 nm and about 5 nm.

44. The method as claimed in any of claim 22 to 42, wherein the ultra-thin metal layer is deposited in step c) by sputtering metal onto the second dielectric layer.

45. The method as claimed in claim 44, wherein in step c) the metal is sputtered for a sputtering period of between about 10 sec and about 60 sec.

46. The method as claimed in claim 44 or 45, wherein in step c) the metal is sputtered for a sputtering period of about 30 sec.

47. The method as claimed in any of claims 22 to 46, wherein the third dielectric layer is deposited in step e) with a thickness of between about 8 nm and about 10 nm.

48. The method as claimed in any of claims 22 to 47, wherein the third dielectric layer is deposited in step e) by atomic layer deposition of the first high-k dielectric material onto the second dielectric layer and the metallic nanocrystals.

49. The method as claimed in any of claims 22 to 48, wherein the first dielectric layer in step a) is provided on a substrate comprising silicon.

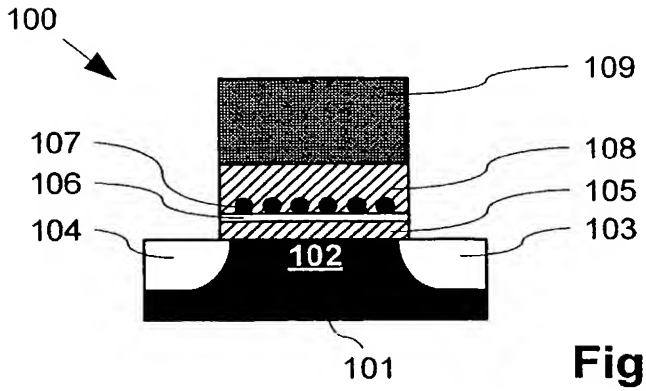

100
109
107
108
106
105
104
103
102
101

Fig.1

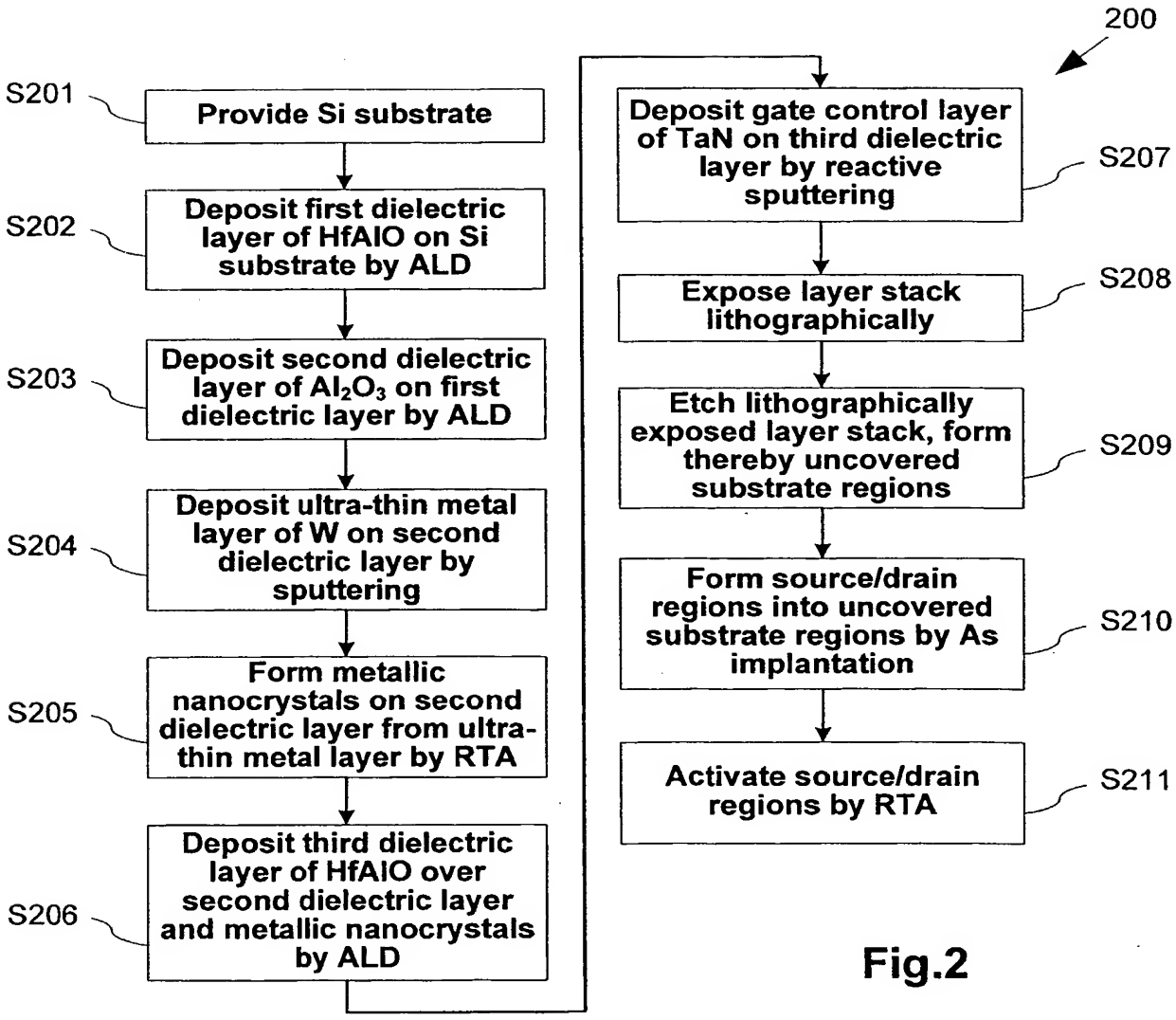

200
S201
Provide Si substrate
S202
Deposit first dielectric layer of HfAlO on Si substrate by ALD
S203
Deposit second dielectric layer of Al_2O_3 on first dielectric layer by ALD
S204
Deposit ultra-thin metal layer of W on second dielectric layer by sputtering
S205
Form metallic nanocrystals on second dielectric layer from ultra-thin metal layer by RTA
S206
Deposit third dielectric layer of HfAlO over second dielectric layer and metallic nanocrystals by ALD
S207
Deposit gate control layer of TaN on third dielectric layer by reactive sputtering
S208
Expose layer stack lithographically
S209
Etch lithographically exposed layer stack, form thereby uncovered substrate regions
S210
Form source/drain regions into uncovered substrate regions by As implantation
S211
Activate source/drain regions by RTA

Fig.2

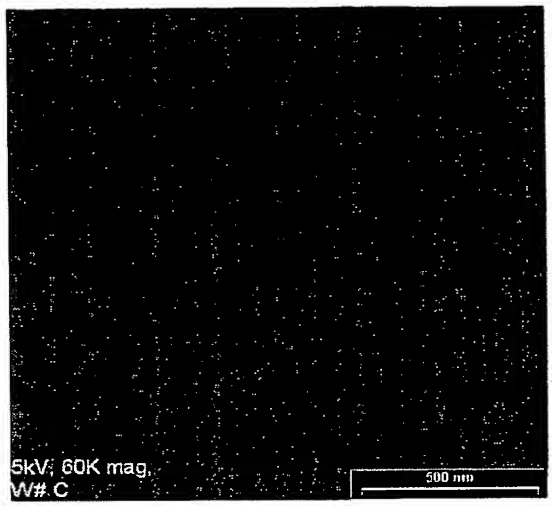


Fig.3A

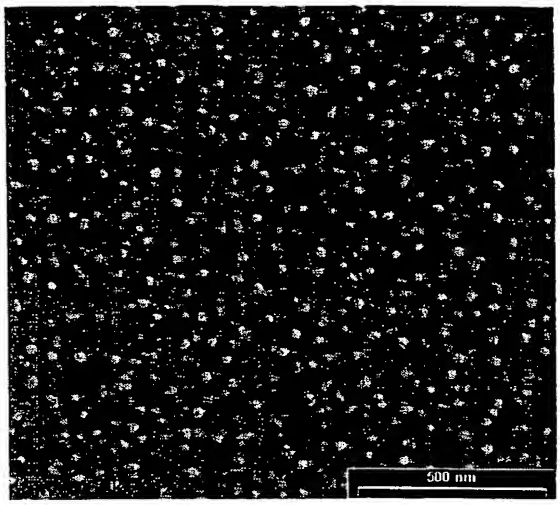


Fig.3B

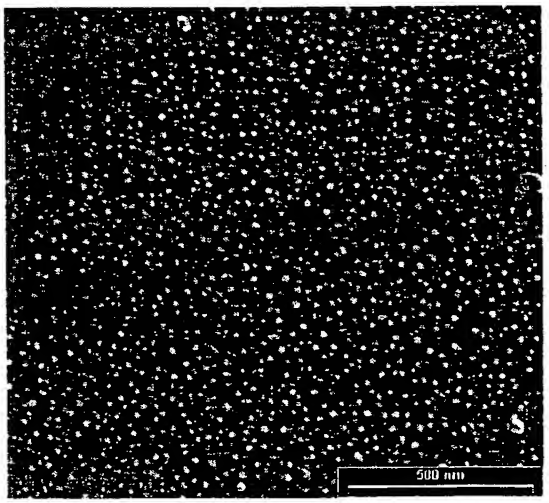


Fig.3C

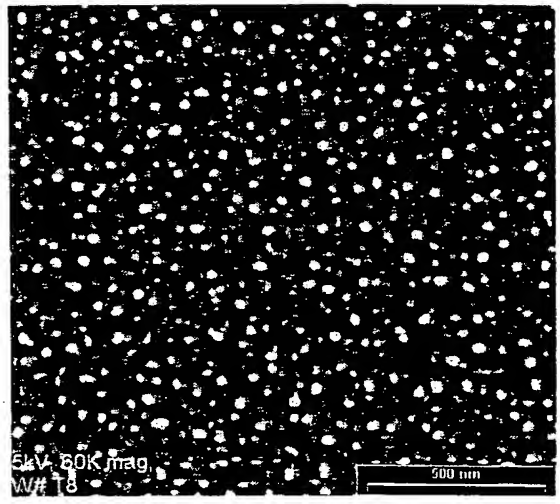


Fig.3D

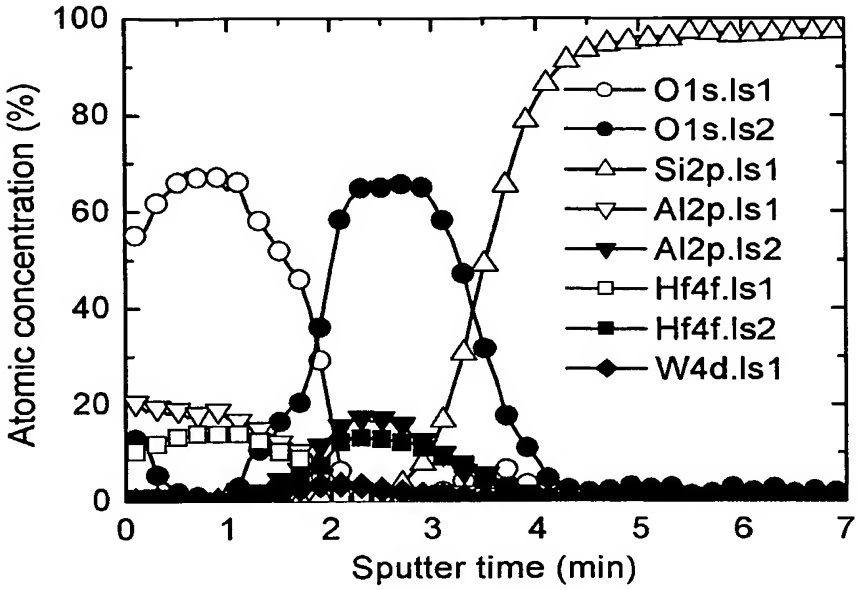

Atomic concentration (%)
Sputter time (min)
O1s.ls1
O1s.ls2
Si2p.ls1
Al2p.ls1
Al2p.ls2
Hf4f.ls1
Hf4f.ls2
W4d.ls1

Fig.4A

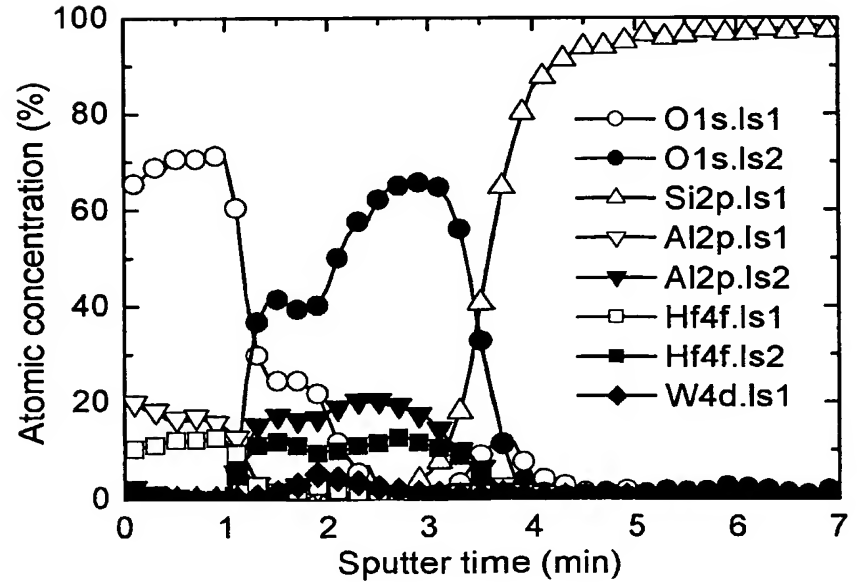

Atomic concentration (%)
Sputter time (min)
O1s.ls1
O1s.ls2
Si2p.ls1
Al2p.ls1
Al2p.ls2
Hf4f.ls1
Hf4f.ls2
W4d.ls1

Fig.4B

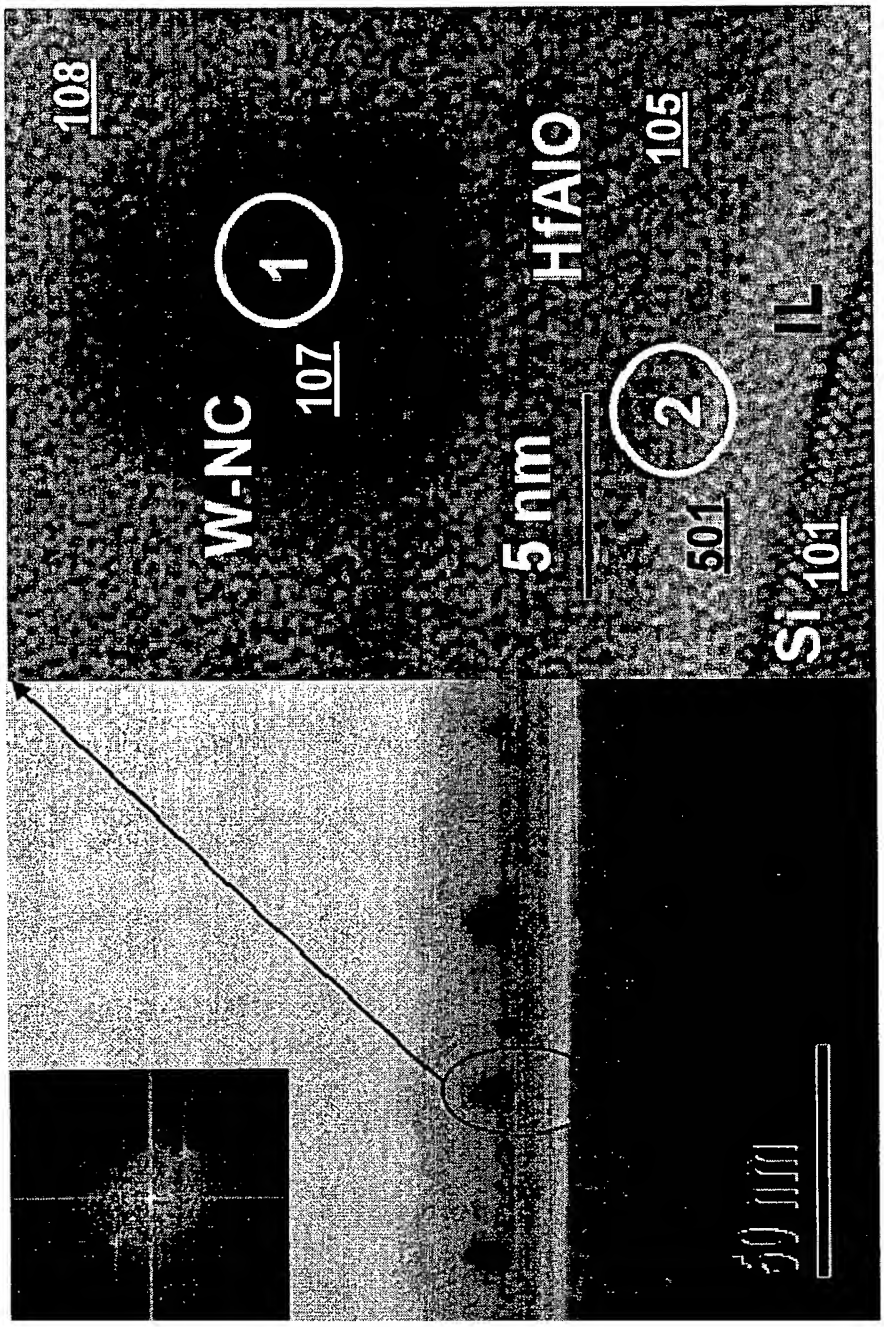


Fig.5B

Fig.5A

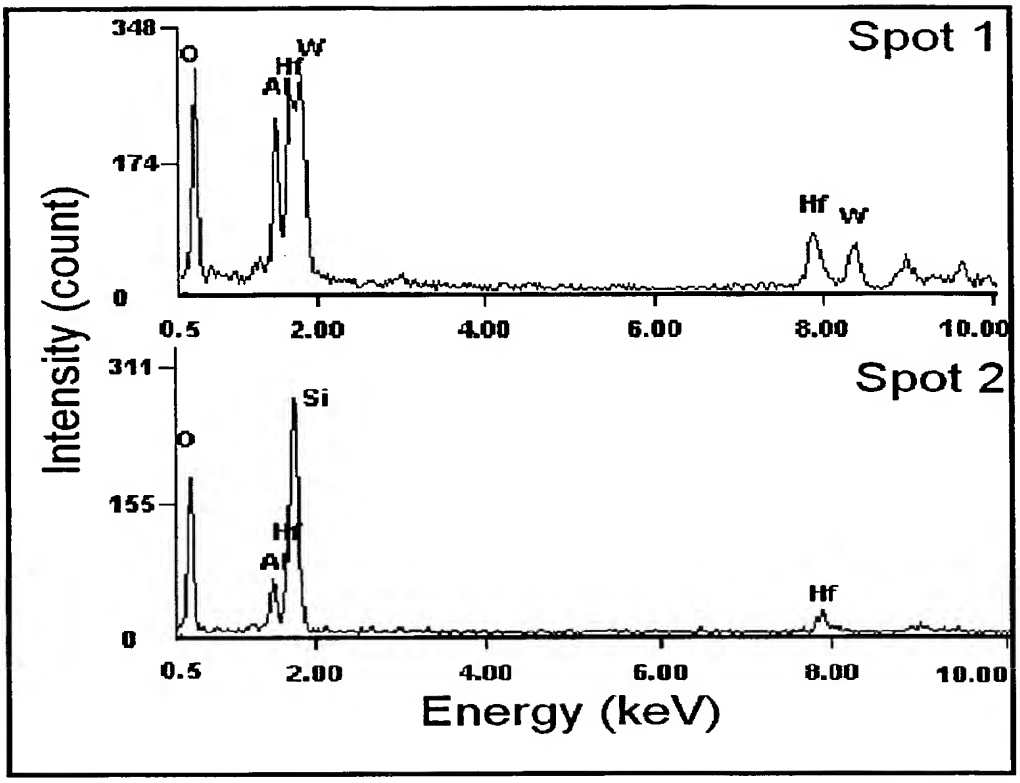


Fig.6A

Element	Spot 1		Spot 2	
	Weight-%	Atomic-%	Weight-%	Atomic-%
Al	12.90	31.90	9.80	11.50
Hf	43.00	16.00	30.70	5.40
O	9.50	39.50	19.50	38.40
Si	--	--	39.90	44.80
W	34.60	12.50	--	--

Fig.6B

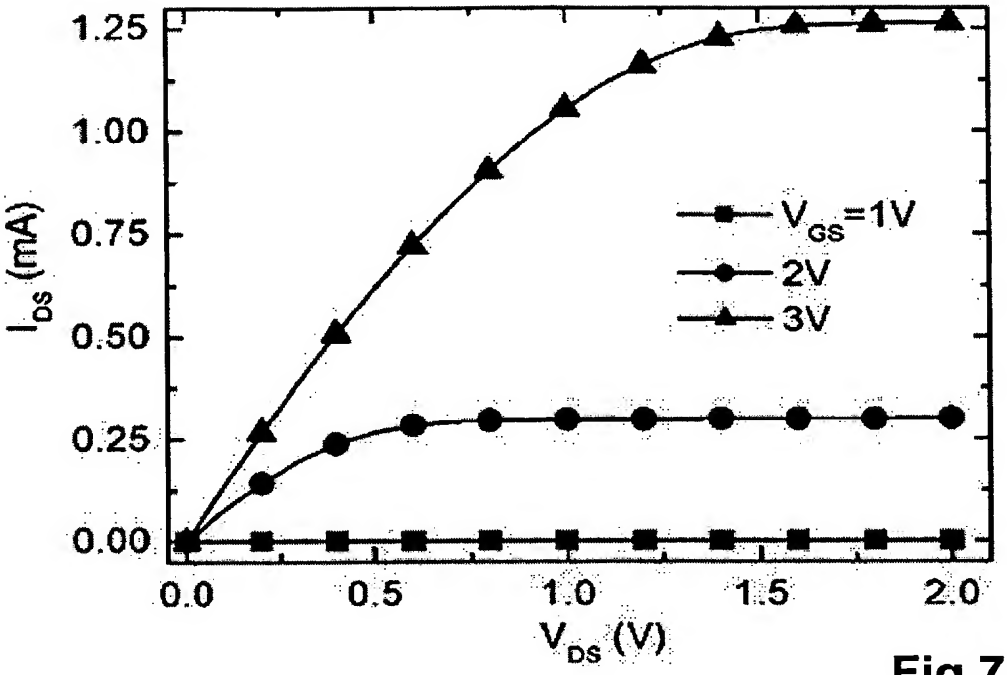

1.25
1.00
0.75
0.50
0.25
0.00
I_{DS} (mA)
0.0
0.5
1.0
1.5
2.0
V_{DS} (V)
V_{GS}=1V
2V
3V

Fig.7

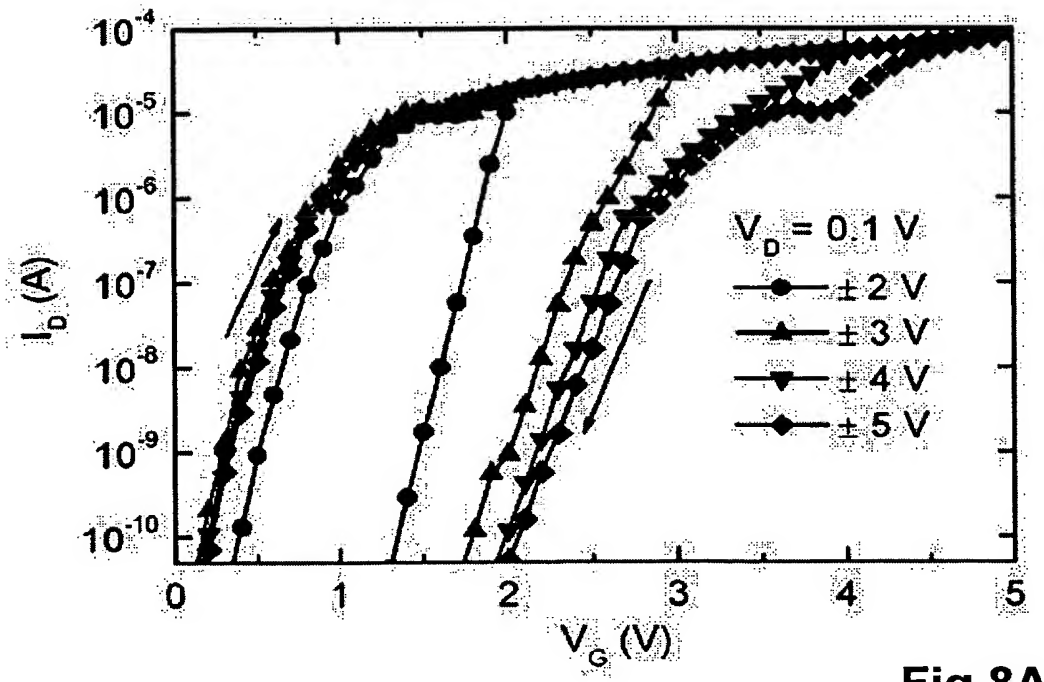

10^{-4}
10^{-5}
10^{-6}
10^{-7}
10^{-8}
10^{-9}
10^{-10}
I_D (A)
0
1
2
3
4
5
V_G (V)
V_D = 0.1 V
± 2 V
± 3 V
± 4 V
± 5 V

Fig.8A

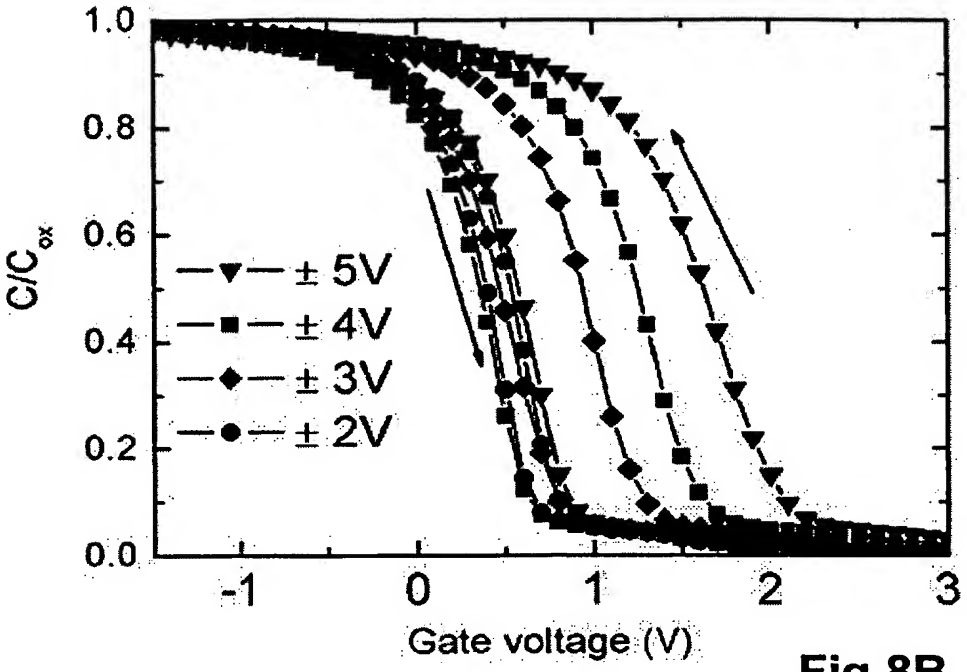

1.0
0.8
0.6
0.4
0.2
0.0
C/C$_{ox}$
± 5V
± 4V
± 3V
± 2V
-1
0
1
2
3
Gate voltage (V)

Fig.8B

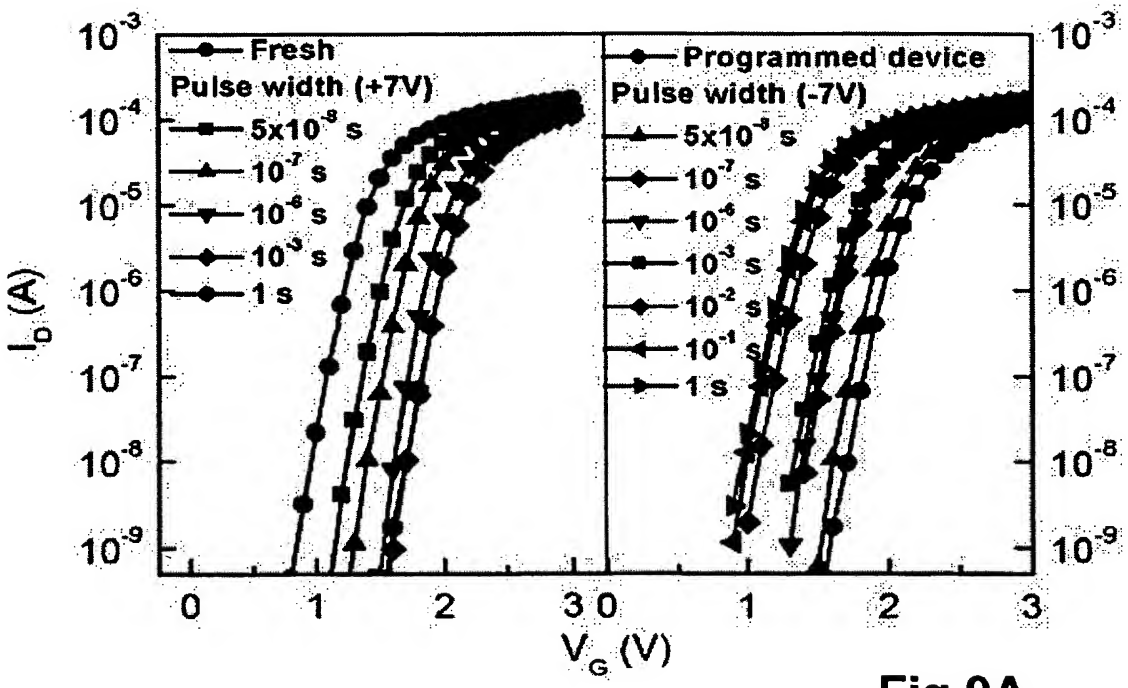

Fresh
Pulse width (+7V)
5x10^{-8} s
10^{-7} s
10^{-6} s
10^{-3} s
1 s
Programmed device
Pulse width (-7V)
5x10^{-8} s
10^{-7} s
10^{-6} s
10^{-3} s
10^{-2} s
10^{-1} s
1 s
I$_D$ (A)
10^{-3}
10^{-4}
10^{-5}
10^{-6}
10^{-7}
10^{-8}
10^{-9}
0
1
2
3
0
1
2
3
V$_G$ (V)

Fig.9A

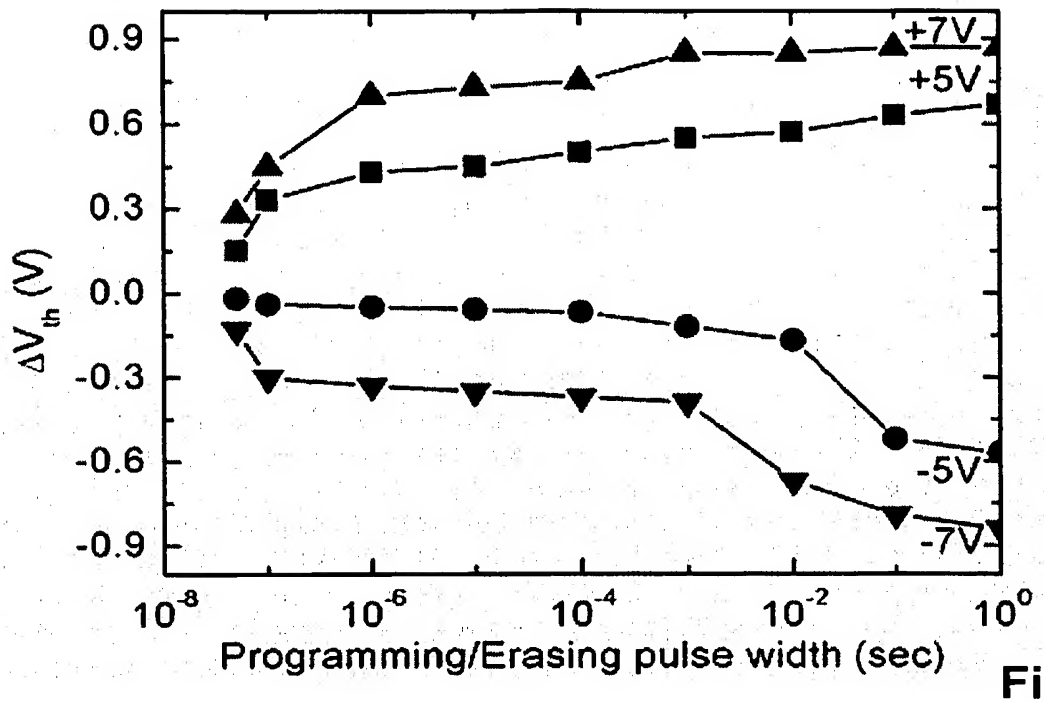

0.9
0.6
0.3
0.0
-0.3
-0.6
-0.9
ΔV_th (V)
+7V
+5V
-5V
-7V
10^-8
10^-6
10^-4
10^-2
10^0
Programming/Erasing pulse width (sec)

Fig.9B

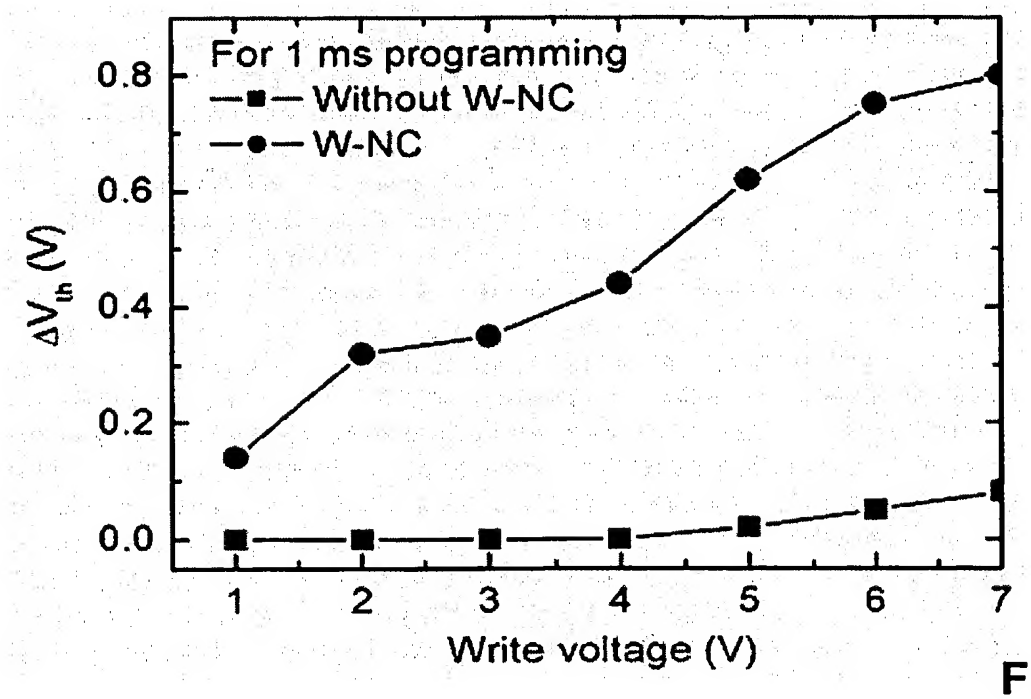

For 1 ms programming
Without W-NC
W-NC
0.8
0.6
0.4
0.2
0.0
ΔV_th (V)
1
2
3
4
5
6
7
Write voltage (V)

Fig.10

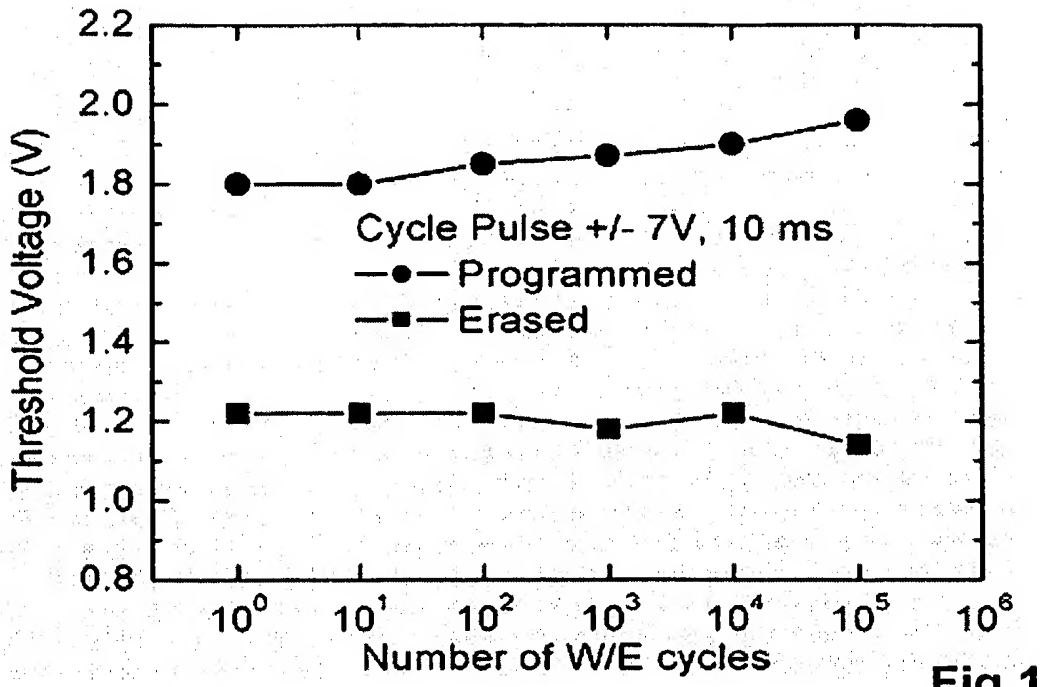

Threshold Voltage (V)
2.2
2.0
1.8
1.6
1.4
1.2
1.0
0.8
Cycle Pulse +/- 7V, 10 ms
Programmed
Erased
10^0 10^1 10^2 10^3 10^4 10^5 10^6
Number of W/E cycles

Fig.11

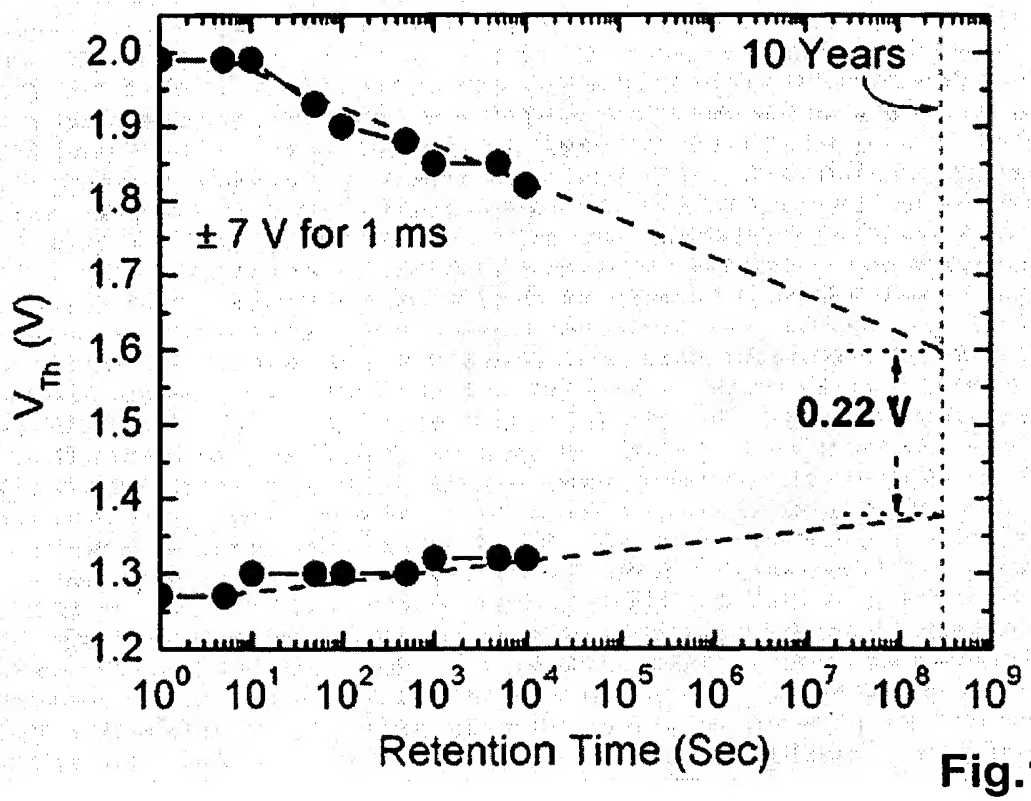

V_{Tn} (V)
2.0
1.9
1.8
1.7
1.6
1.5
1.4
1.3
1.2
10 Years
±7 V for 1 ms
0.22 V
10^0 10^1 10^2 10^3 10^4 10^5 10^6 10^7 10^8 10^9
Retention Time (Sec)

Fig.12

Results of	Material of metallic nanocrystals	Material of tunneling oxide	Thickness of tunneling oxide	Material of control oxide	Thickness of control oxide	V_{th} decay rate (mV/dec)
present applicatioin	W	HfAlO	6 nm	HfAlO	10 nm	50
prior art	W	SiO_2	5 nm	SiO_2	8 nm	250
prior art	Au (double layer)	SiO_2	2-3 nm	SiO_2	36 nm	125
prior art	Au (single layer)	SiO_2	2-3 nm	SiO_2	36 nm	175
prior art	Pt	SiO_2	8 nm	SiO_2	30 nm	20
prior art	Ag	SiO_2	8 nm	SiO_2	30 nm	25
prior art	Au	SiO_2	8 nm	SiO_2	30 nm	75

Fig.13

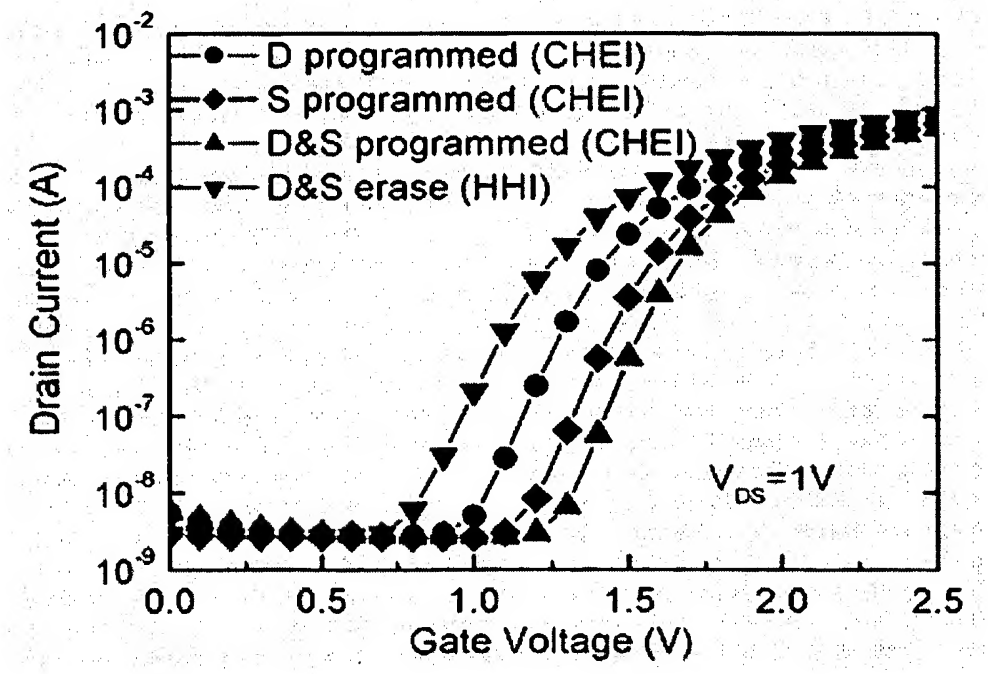


Fig.14

INTERNATIONAL SEARCH REPORT

International application No.
PCT/SG2005/000129

A. CLASSIFICATION OF SUBJECT MATTER

Int. Cl. [7]: G11C 11/34

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
WPAT, PCT Gazette, IEEE Xplore, USPTO Web Patent Database "memory, nanocrystals, metal, dielectric etc."

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 2003/0235064 A1 (BATRA et al.) 25 December 2003 Whole document.	1-49
X	Chen et al. "Nonvolatile Flash Memory Device Using Ge Nanocrystals Embedded In HfAlO High-k Tunneling and Control Oxides: Device Fabrication and Electrical Performance," IEEE Transactions on Electron Devices, vol. 51, no. 11, Nov. 2004 Pages 1840 to 1848.	1,9-11,16-21
Y	Pages 1840 to 1848 in combination with any of the other documents.	1,9-11,15-21
Y	Lee et al. "Metal Nanocrystal Memory with High-k Tunneling Barrier for Improved Data Retention", IEEE Transactions on Electron Devices, vol. 52, no. 4, Apr. 2005 Pages 507 to 511 in combination with any of the other documents.	1,9-11,15-21
Y	She et al. " Impact of crystal size and tunnel dielectric on semiconductor nanocrystal memory performance," IEEE Transactions on Electron Devices, vol 50, no. 9, Sep. 2003 Pages 1934 to 1940 in combination with the above documents.	1,9-11,15-21

☐ Further documents are listed in the continuation of Box C ☒ See patent family annex

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
1 July 2005	11 JUL 2005
Name and mailing address of the ISA/AU AUSTRALIAN PATENT OFFICE PO BOX 200, WODEN ACT 2606, AUSTRALIA E-mail address: pct@ipaustralia.gov.au Facsimile No. (02) 6285 3929	Authorized officer **P. THONG** Telephone No : (02) 6283 2128

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No.
PCT/SG2005/000129

This Annex lists the known "A" publication level patent family members relating to the patent documents cited in the above-mentioned international search report. The Australian Patent Office is in no way liable for these particulars which are merely given for the purpose of information.

Patent Document Cited in Search Report		Patent Family Member
US	2003/0235064	NONE

Due to data integration issues this family listing may not include 10 digit Australian applications filed since May 2001.

END OF ANNEX